摘要

随着网络视频业务与智能多媒体应用的迅猛发展，视频数据规模持续攀升，带宽与存储资源压力日益加剧，推动视频压缩技术向更高效率与更强智能化方向演进。现有视频编码方法，无论是传统标准体系还是深度学习驱动的端到端模型，大多基于统一的率失真优化策略，对不同空间区域采用一致的编码方式，主要依赖统计冗余压缩来提升编码效率。然而，这类方法忽略了视频内容在语义层面与视觉感知层面的不均匀性，在低码率条件下容易造成关键目标区域细节缺失，而背景区域却占用过多比特资源，从而限制了整体压缩性能的进一步提升。

针对上述问题，本文围绕神经网络视频压缩方法的高质量视觉重建展开研究，提出一种面向感知质量优化的端到端视频编码方法，通过引入显著性建模机制，对视频中的重要区域进行优先建模与编码，能够在有限码率预算下实现更合理的比特分配，从而在兼顾编码效率的前提下，提升重建视频的任务相关性与感知质量。主要创新点包括以下两个方面：

第一，提出面向智能视频编码的显著性区域提取方法。针对通用显著性检测方法与视频压缩目标之间存在偏差的问题，本文从编码优化角度重新定义显著性区域的判别标准，将语义重要性、视觉感知敏感性以及重建误差分布进行联合建模。具体而言，通过引入语义分割或目标检测网络获取高层语义先验，并结合编码重建过程中的失真反馈信息，对不同空间位置的重要性进行自适应评估，从而构建与率失真优化目标一致的显著性表示。在模型设计上，进一步融合时序信息以刻画视频帧间的动态变化特性，使显著性区域不仅在空间维度上具有准确性，同时在时间维度上保持一致性与稳定性。此外，为提升显著性提取与编码任务之间的协同优化能力，本文设计端到端联合训练策略，使显著性预测模块能够在压缩目标的直接约束下进行参数更新，从而获得更加贴合编码需求的显著性掩码表示。实验结果表明，该方法相比传统显著性检测或简单启发式区域划分策略，在关键区域定位准确性与编码任务相关性方面均具有明显优势，为后续差异化编码提供了可靠且有效的先验信息。

第二，提出基于感兴趣区域（Region of Interest, ROI）引导的视频压缩优化方法。针对传统统一编码策略难以适应视频内容区域重要性差异、导致码率分配效率不足的问题，目前在神经网络视频编码（NVC）领域尚未有工作系统性地引入 ROI 感知编码机制，尚缺乏充分利用区域重要性差异的统一优化框架。为此，创新性地提出一种面向感兴趣区域优化的神经网络视频压缩方法，对关键视觉区域进行建模与自适应增强。通过引入区域掩码（ROI mask）作为先验信息，引导

编码过程进行区域感知建模。此外，通过在特征层面对重要区域进行重标定与调制，使编码过程能够根据语义重要性动态调整信息分配策略，从而优先保障关键区域的表示质量，实现结构信息保真与感知质量提升的协同优化，在相同比特率约束下获得更优的主观视觉效果和重建质量。随后在潜变量空间中引入自适应缩放机制，根据区域重要性对潜变量表示进行动态调制，以实现更精细的比特资源分配。实验结果表明，该方法在不显著增加模型复杂度的前提下，在多个数据集上能够显著提升感兴趣区域重建质量与整体视觉感知效果，验证了在神经网络视频编码框架中引入感兴趣区域编码的有效性与必要性。

综上所述，本文围绕神经网络视频编码中的主观质量优化问题，提出既考虑人眼视觉效果又考虑率失真性能的神经网络视频压缩方法，从显著性区域建模与局部区域感知增强两个层面实现协同优化。所提出方法提升了视频压缩的感知质量与编码效率，为高质量视频重建提供了一种有效途径，并为神经网络视频编码技术的发展提供了新的研究思路。